The Largest Nonidentifiable Outlier: A Comparison of Multivariate Simultaneous Outlier Identification Rules

Claudia Becker and Ursula Gather

Department of Statistics, University of Dortmund, D–44221 Dortmund, Germany

Abstract

The aim of detecting outliers in a multivariate sample can be pursued in different ways. We investigate here the performance of several simultaneous multivariate outlier identification rules based on robust estimators of location and scale. It has been shown that the use of estimators with high finite-sample breakdown point in such procedures yields a good behaviour with respect to the prevention of breakdown by the masking effect (Becker, Gather, 1999, J. Amer. Statist. Assoc. 94, 947-955). In this article, we investigate by simulation, at which distance from the center of an underlying model distribution outliers can be placed until certain simultaneous identification rules will detect them as outliers. We consider identification procedures based on the minimum volume ellipsoid, the minimum covariance determinant, and S-estimators.

Keywords: Outliers; high breakdown point procedures; MVE; MCD; robustness; S-estimators

1 Introduction

The use of robust estimators for detecting outliers in multivariate data simultaneously has been proposed by several authors (e.g. Simonoff, 1987). Becker and Gather (1999) give a formal justification for this, showing that using robust estimators with high finite-sample breakdown point leads to better prevention from the masking effect than using classical estimators in such simultaneous outlier identification rules. Several multivariate location and scale estimators can therefore be taken into account, for example the MVE- and MCD-estimators and reweighted versions of them (Rousseeuw, 1985, Lopuhaä, Rousseeuw, 1991, Croux, Haesbroeck, 2000), constrained M-estimators (Kent, Tyler, 1996), various types of S-estimators (Davies, 1987, Maronna, Yohai, 1995, Rocke, 1996), or the Stahel-Donoho estimators (Stahel, 1981, Donoho, 1982, Tyler, 1994, Maronna, Yohai, 1995, Gather, Hilker, 1997).

We restrict ourselves to the case of the p-variate normal $N(\underline{\mu}, \Sigma)$, $\underline{\mu} \in I\!R^p$, $S \in I\!R^{p\times p}$ positive definite (p.d.), as model distribution. Following the idea of Davies and Gather (1993), we then consider the aim of detecting all α_N outliers in a sample of size N, i.e. all observations lying in the α_N outlier region $\mathrm{out}(\alpha_N, \underline{\mu}, \Sigma)$ of $N(\underline{\mu}, \Sigma)$:

$$\mathrm{out}(\alpha_N, \underline{\mu}, \Sigma) = \{\underline{x} \in I\!R^p : (\underline{x} - \underline{\mu})^T \Sigma^{-1} (\underline{x} - \underline{\mu}) > \chi^2_{p;1-\alpha_N}\},$$

where $\alpha_N = 1 - (1-\alpha)^{1/N}$ for some given value of $\alpha \in (0,1)$, such that

$$P_{N(\underline{\mu},\Sigma)}(\text{no observation in a sample of size } N \text{ lies in } \mathrm{out}(\alpha_N, \underline{\mu}, \Sigma)) = 1 - \alpha.$$

Taking usual choices for α ($\alpha = 0.05, 0.1$), this means that under the model distribution there is only a small probability for any observation of a sample of size N to lie in the outlier region, reflecting the intuitive idea that an outlier is an "unexpected" and far out observation.

Let now $(\underline{m}, S)$ be a pair of affine equivariant estimators for $(\underline{\mu}, \Sigma)$. To find all α_N outliers in a given sample $\underset{\sim}{x}_N = (\underline{x}_1, \ldots, \underline{x}_N)$ of size N, $\underline{x}_i \in I\!R^p$, we can use $(\underline{m}, S)$ to estimate out$(\alpha_N, \underline{\mu}, \Sigma)$ by a corresponding region $\underline{\mathrm{OR}}(\underset{\sim}{x}_N, \alpha_N)$, called outlier identifier,

$$\underline{\mathrm{OR}}(\underset{\sim}{x}_N, \alpha_N) = \{\underline{x} \in I\!R^p : (\underline{x} - \underline{m})^T S^{-1} (\underline{x} - \underline{m}) \geq c(p, N, \alpha_N)\},$$

where $c(p, N, \alpha_N)$ is a suitably chosen constant, calculated according to some normalizing condition, for example

$$P_{N(\underline{\mu},\Sigma)}(\text{no } \alpha_N \text{ outliers identified in a sample of size } N) = 1 - \alpha \tag{1.1}$$

with $\alpha_N = 1 - (1 - \alpha)^{1/N}$ as before. This is equivalent to keeping a level α when applying consecutive testing methods to identify outliers (Davies, Gather, 1993). An observation $\underline{x}_i$ lying in $\underline{\mathrm{OR}}(\underset{\sim}{x}_N, \alpha_N)$ is then detected as an outlier. The use of affine equivariant estimators guarantees that the identification rule is affine equivariant, too. For more details on the concept of α_N outliers see Davies, Gather (1993), Gather, Becker (1997), or Becker, Gather (1999).

The simultaneous outlier identification procedure described above corresponds to calculating the Mahalanobis-type distances $d_i^2 = (\underline{x}_i - \underline{m})^T S^{-1} (\underline{x}_i - \underline{m})$, $i = 1, \ldots, N$, with respect to $(\underline{m}, S)$ and identifying an observation with large d_i^2 $(\geq c(p, N, \alpha_N))$ as an outlier (see e.g. Barnett, Lewis, 1994, pp. 306ff., Rousseeuw, van Zomeren, 1990, for procedures of this kind).

In this paper, we compare outlier identifiers $\underline{\mathrm{OR}}$ based on some of the above mentioned robust estimators with respect to a certain performance criterion, namely the size of the largest nonidentifiable outlier. The question is how "far away" an observation can be while still not being detected as an outlier.

Our paper is organized as follows. In Section 2, we present the robust estimators $\underline{m}$ and S used in the identification rules. In Section 3, we discuss which positions of outliers represent a worst case situation for these procedures and introduce the notion of the "largest nonidentifiable outlier". We then give the results of a simulation study, comparing the behaviour of the identifiers with respect to this criterion.

2 Outlier Identifiers Based on Robust Estimators

We will discuss the behaviour of three multivariate outlier identifiers based on the following robust estimators of location and scatter: the MVE- and MCD-estimators of Rousseeuw (1985), and a pair $(\underline{m}, S)$ of S-estimators using Tukey's biweight function according to Rocke (1996). These estimators are of similar type: they minimize the volume of some ellipsoid. We focus on this class of estimators to investigate whether there is a "best" choice among them. Moreover, there exist feasible algorithms to calculate them in moderate computation time (Rocke, 1996, Rousseeuw and van Driessen, 1999, Rousseeuw and van Zomeren, 1990). In further investigations, it will be interesting to compare the best identifier found here with an identifier based on robust projection pursuit estimators like the Stahel-Donoho estimators (Donoho, 1981, Maronna and Yohai, 1995, Stahel, 1982, Tyler, 1994).

All estimators considered are chosen such that they have the maximum possible finite-sample breakdown point in the sense of Donoho and Huber (1983). We use the definition of the finite-sample breakdown point of an estimator as the smallest proportion of the data that needs to be replaced by arbitrary points to cause the breakdown of the estimate.

For a location estimator, breakdown occurs if the estimate becomes arbitrarily large in the sense of an infinite euclidean distance. In the case of a multivariate scatter estimate, we speak of breakdown if either the smallest eigenvalue of the estimated matrix becomes arbitrarily close to zero or if its largest eigenvalue grows beyond all bounds. See Donoho, Huber (1983) and Lopuhaä, Rousseeuw (1991) for formal definitions. The maximum possible finite-sample breakdown point for pairs $(\underline{m}, S)$ of affine equivariant estimators of location and covariance is $[(N-p+1)/2]/N$ (Davies, 1987), $[x]$ denoting the integer part of $x \in \mathbb{R}$. Using such estimators for outlier detection also means that the resulting outlier identifiers give the best possible protection against the masking effect (Becker, Gather, 1999). Roughly spoken, masking occurs if the detection of outlying observations is made impossible by the presence of some extreme outliers in the data. In this sense, each of the estimators considered here can be seen as an optimal choice to be used in a simultaneous outlier identification procedure.

The Identifiers $\underline{\mathbf{OR}}_{\mathrm{MVE}}$ and $\underline{\mathbf{OR}}_{\mathrm{MCD}}$

The multivariate outlier identifier based on the MVE-estimators is defined as

$$\underline{\mathrm{OR}}_{\mathrm{MVE}} := \{\underline{x} \in \mathbb{R}^p : (\underline{x} - \underline{m}_{\mathrm{MVE}})^T S_{\mathrm{MVE}}^{-1} (\underline{x} - \underline{m}_{\mathrm{MVE}}) \geq c_{\mathrm{MVE}}(p, N, \alpha_N)\},$$

where $\underline{m}_{\mathrm{MVE}}$ is the center of the minimum volume ellipsoid (MVE) covering at least $h = [(N+p+1)/2]$ of the data $\underset{\sim}{x}_N$, and S_{MVE} is the sample covariance matrix of the data points lying within this MVE, multiplied by a constant factor to obtain Fisher consistency for the multivariate normal distribution. For the identifier $\underline{\mathrm{OR}}_{\mathrm{MCD}}$, we use $\underline{m}_{\mathrm{MCD}}$ and S_{MCD}, the mean and (normalized) sample covariance of a certain subset of the data. This subset consists of those h points of $\underset{\sim}{x}_N$ yielding the smallest determinant of the covariance matrix (minimum covariance determinant, MCD). As for S_{MVE}, we have to multiply the covariance matrix with a constant to obtain the consistent estimator S_{MCD} (Rousseeuw, van Driessen, 1999). The choice of h guarantees that both pairs of estimators have maximum breakdown points (Davies, 1987, Lopuhaä, Rousseeuw, 1991).

Table 1. Normalizing constants for $\alpha = 0.1$ compared to the respective χ^2 quantiles

N	p	$\chi^2_{p;1-\alpha_N}$	c_{MVE}	c_{MCD}	c_{BW}
20	2	10.49746	19.14222	85.58786	21.35944
20	3	12.73173	23.47072	167.61310	26.87044
20	4	14.74769	33.72110	388.84680	33.17018
50	2	12.32689	17.54896	28.51695	16.93195
50	3	14.68664	20.61580	41.83594	19.78682
50	4	16.80930	24.65417	64.18462	23.14061

The constants c_{MVE} and c_{MCD} are calculated according to the normalizing condition (1.1) by simulation using the algorithms of Rousseeuw and van Driessen (1999) and Rousseeuw and van Zomeren (1990), as implemented in the statistical package S-Plus (version 4.5). In work on the identification of outliers, these normalizing constants are often chosen to be quantiles of the χ^2_p-distribution, using the asymptotics of the Mahalanobis-type distances $(\underline{x} - \underline{m})^T S^{-1} (\underline{x} - \underline{m})$. We do not follow this approach, because the approximation turns out to be rather bad for the cases considered here. In our simulation study, we will look at samples of size $N = 20, 50$ in dimension $p = 2, 3, 4$. Using the χ^2 approximation would mean to take $c(p, N, \alpha_N) = \chi^2_{p;1-\alpha_N}$ with $\alpha_N = 1 - (1-\alpha)^{1/N}$ as before. In Table 1, we give the values of $\chi^2_{p;1-\alpha_N}$ compared to the values of $c(p, N, \alpha_N)$ for $\underline{\mathrm{OR}}_{\mathrm{MVE}}$ and $\underline{\mathrm{OR}}_{\mathrm{MCD}}$ calculated from 10000 observations. It is obvious that the χ^2 approximation is not appropriate here.

The Identifier $\underline{\mathbf{OR}}_{\mathrm{BW}}$

We discuss a further outlier identifier, $\underline{\mathrm{OR}}_{\mathrm{BW}}$, which is based on S-estimators that are constructed using Tukey's biweight function (BW; Beaton, Tukey, 1974). The pair $(\underline{m}_{\mathrm{BW}}, S_{\mathrm{BW}})$ of estimators is found by solving the following general minimization problem (Lopuhaä, 1989):

$$\min_{S \,\in\, PDS(p)} \det(S),$$

under the restriction

$$1/N \sum_{i=1}^{N} \rho\left(\sqrt{(\underline{X}_i - \underline{m})^T S^{-1} (\underline{X}_i - \underline{m})}\right) \;=\; b_0,$$

where we choose ρ to be the special function $\rho_{\mathrm{BW}} : I\!R_+ \mapsto I\!R$ with

$$\rho_{\mathrm{BW}}(d) = \begin{cases} d^2/2 - d^4/(2c_0^2) + d^6/(6c_0^4) & , \quad 0 \le d \le c_0, \\ \\ c_0^2/6 & , \quad d > c_0. \end{cases}$$

The derivative of ρ_{BW} is known as Tukey's biweight function; $PDS(p)$ denotes the set of positive definite symmetric $p \times p$ matrices. The constants b_0 and c_0 are determined such that $\underline{m}_{\mathrm{BW}}$ and S_{BW} possess maximum breakdown points. That means, c_0 solves $\mathrm{E}(\rho_{\mathrm{BW}}(D)) = ([(N-p+1)/2]\,/N)\rho_{\mathrm{BW}}(c_0)$, where D is a random variable with $D^2 \sim \chi^2_p$. The constant b_0 is calculated from $\mathrm{E}(\rho_{\mathrm{BW}}(D)) = b_0$, where the expectation is taken with respect to the multivariate normal. In Rocke (1996), an iteration scheme is given to obtain $\underline{m}_{\mathrm{BW}}$ and S_{BW}. As before, the normalizing constant c_{BW} is calculated by simulation according to (1.1). The results can also be found in Table 1. For all simulations we choose a value of $\alpha = 0.1$.

Rocke (1996) discusses the behaviour of the biweight S-estimators. He finds that especially in higher dimensions the influence of large outliers on such estimators can be rather strong. For this reason he develops a modification ρ_{TW} of ρ_{BW} called translated biweight. We will concentrate on samples in moderate dimension here, thus we will not investigate the estimators resulting from ρ_{TW} in the following. Trials with some selected simulation constellations show that the respective identifier $\underline{\mathrm{OR}}_{\mathrm{TW}}$ indeed cannot compete with the other three identification procedures in our case. The case of higher dimensions is still under research and will be treated elsewhere.

3 The Largest Nonidentifiable Outlier

To investigate the performance of the identification procedures defined above, without loss of generality we set $\underline{\mu} = \underline{0}$ and $\Sigma = \mathcal{I}$ because of the affine equivariance of all estimators considered here. Thus, we can look at the size of an observation (with respect to $N(\underline{0}, \mathcal{I})$) as its euclidean distance from the origin, and we can ask which size an outlier may have, while still not being detected by the above outlier identifiers. We assume that the proportion of outliers in the data does not exceed the finite-sample breakdown point of the estimators used in the identifiers. This means that the number k of outliers in the data is smaller than $[(N - p + 1)/2]$.

For the estimators considered here, the worst-case situation is naturally given when concentrating all outliers at one point and placing them in a certain distance from the origin. This leads to the idea of taking the size of the largest nonidentifiable outlier as an interesting worst case performance criterion of an outlier identifier. Let $\underset{\sim}{x}_N = (\underline{x}_1, \ldots, \underline{x}_N)$ be a sample of size N from $N(\underline{0}, \mathcal{I})$, $\underset{\sim}{x}_k^0 = (\underline{y}_1, \ldots, \underline{y}_k)$ with $\underline{y}_i \in$ out$(\delta_N, \underline{0}, \mathcal{I})$ for some suitable δ_N. Replace k of the observations of $\underset{\sim}{x}_N$ by $\underset{\sim}{x}_k^0$, yielding $\underset{\sim}{x}_{N,k} = (\underline{x}_{i_1}, \ldots, \underline{x}_{i_n}, \underline{y}_1, \ldots, \underline{y}_k)$, where $n = N - k$. Define the size of the largest nonidentifiable outlier as

$$\sup_{\underset{\sim}{x}_k^0 \in \mathrm{out}(\delta_N, \underline{0}, \mathcal{I})} \{\|\underline{y}\| : \underline{y} \in \underset{\sim}{x}_k^0 \text{ and } (\underline{y} - \underline{m}(\underset{\sim}{x}_{N,k}))^T S(\underset{\sim}{x}_{N,k})^{-1}(\underline{y} - \underline{m}(\underset{\sim}{x}_{N,k})) < c(p, N, \alpha_N)\}.$$

From the above considerations, for the identifiers investigated here, it suffices to choose $\underset{\sim}{x}_k^0 = (\underline{y}, \ldots, \underline{y})$, thus the size of the largest nonidentifiable outlier is defined as

$$\sup_{\underline{y} \in \mathrm{out}(\delta_N, \underline{0}, \mathcal{I})} \{\|\underline{y}\| : (\underline{y} - \underline{m}(\underset{\sim}{x}_{N,k}))^T S(\underset{\sim}{x}_{N,k})^{-1}(\underline{y} - \underline{m}(\underset{\sim}{x}_{N,k})) < c(p, N, \alpha_N)\}. \tag{3.1}$$

Theoretical results for the finite-sample performance of the estimators used in the identification rules investigated here are barely known. Available results mainly concern the finite-sample breakdown points of the estimators (e.g. Rousseeuw, 1985) and the masking breakdown points of the outlier identifiers (Becker and Gather, 1999). Simulations are needed to supplement these findings, to get an impression of the behavior of the rules in finite samples. Thus, we calculated the largest nonidentifiable outliers for the three outlier identification rules introduced above for samples of size $N = 20, 50$ and several values of k in a simulation study. For $N = 20$, we considered the cases $k = 1, \ldots, 7$, for $N = 50$, we took $k = 1(2)21$. In each case, we generated samples $\underset{\sim}{x}_{N,k}$ according to the scheme given above and without loss of generality took $\underline{y} = (y_1, 0, \ldots, 0)^T$. We calculated the size of the largest nonidentifiable outlier according to (3.1) for an amount $\eta = k/N$ of outliers and took the mean size from 1000 simulation runs each. The simulated sizes turn out to be relatively stable, without "outliers" occurring in the simulated values themselves. Means and medians of the simulated sizes of the largest nonidentifiable outliers are close together in almost all of the considered constellations. Thus, using the median instead of the mean size leads to the same conclusions. The only case where there is a remarkable difference between mean and median size is for $\underline{\mathrm{OR}}_{\mathrm{MCD}}$ when $N = 50$, $p = 4$, $k = 21$. There, the mean of the simulated values equals 3185.10, whereas the median value is 2741.17. But this does not change the overall result of $\underline{\mathrm{OR}}_{\mathrm{MCD}}$ being the worst of the three identifiers in this case, it only shows that this situation is practically equivalent to a breakdown as is also discussed below. The results for the three identifiers are shown in an overview in Figure 1 and in a magnified version showing more details in Figure 2.

Two main conclusions can be drawn from these results. First, we see that none of the investigated rules allows arbitrarily large outliers to remain undetected, but it still can happen that very large observations are not identified as outliers. To get an impression, take the case of $N = 50$, $p = 2$, $k = 5$ for the identifier $\underline{\mathrm{OR}}_{\mathrm{MVE}}$ which yields the best result for that constellation (see Figure 2). The mean size of the largest nonindentifiable outlier in this case is 4.17. This corresponds to the five points being α_N outliers for a value of $\alpha \approx 0.0083$. Of course, the situation becomes worse with an increasing proportion η of outliers in the sample, the results being worst if η is close to the finite-sample breakdown point of the estimators and, therefore, to the masking breakdown point of the identification procedure (see Becker, Gather, 1999). Even if η remains smaller than the finite-sample breakdown point, we see e.g. that the case of 7 outliers within 20 observations in a four-dimensional space is practically equivalent to a breakdown of the identification procedure. It is also worth noting that with increasing dimension the mean size of the largest nonidentifiable outlier increases drastically.

The second conclusion refers to the behaviour of the different outlier identification rules. From Figure 1, we can see that although none of the rules investigated here performs uniformly optimal, the identifier based on the biweight S-estimators performs quite well in most of the cases. For a small amount of outliers, $\underline{\mathrm{OR}}_{\mathrm{BW}}$ generally behaves best (see Figure 2); for medium numbers of outliers, $\underline{\mathrm{OR}}_{\mathrm{MCD}}$ should be slightly preferred. In smaller samples, $\underline{\mathrm{OR}}_{\mathrm{MVE}}$ yields results similar to $\underline{\mathrm{OR}}_{\mathrm{BW}}$, but for larger samples, again $\underline{\mathrm{OR}}_{\mathrm{BW}}$ performs better. This is also the case, if the dimension of the data is relatively large with respect to the sample size ($N = 20, p = 4$; $N = 50, p = 4$). Altogether, $\underline{\mathrm{OR}}_{\mathrm{BW}}$ leads to the best results in the majority of the cases. Thus, our conclusion is that, for simultaneous identification of outliers in multivariate samples of moderate dimension, the procedure based on the biweight S-estimators should be favorized.

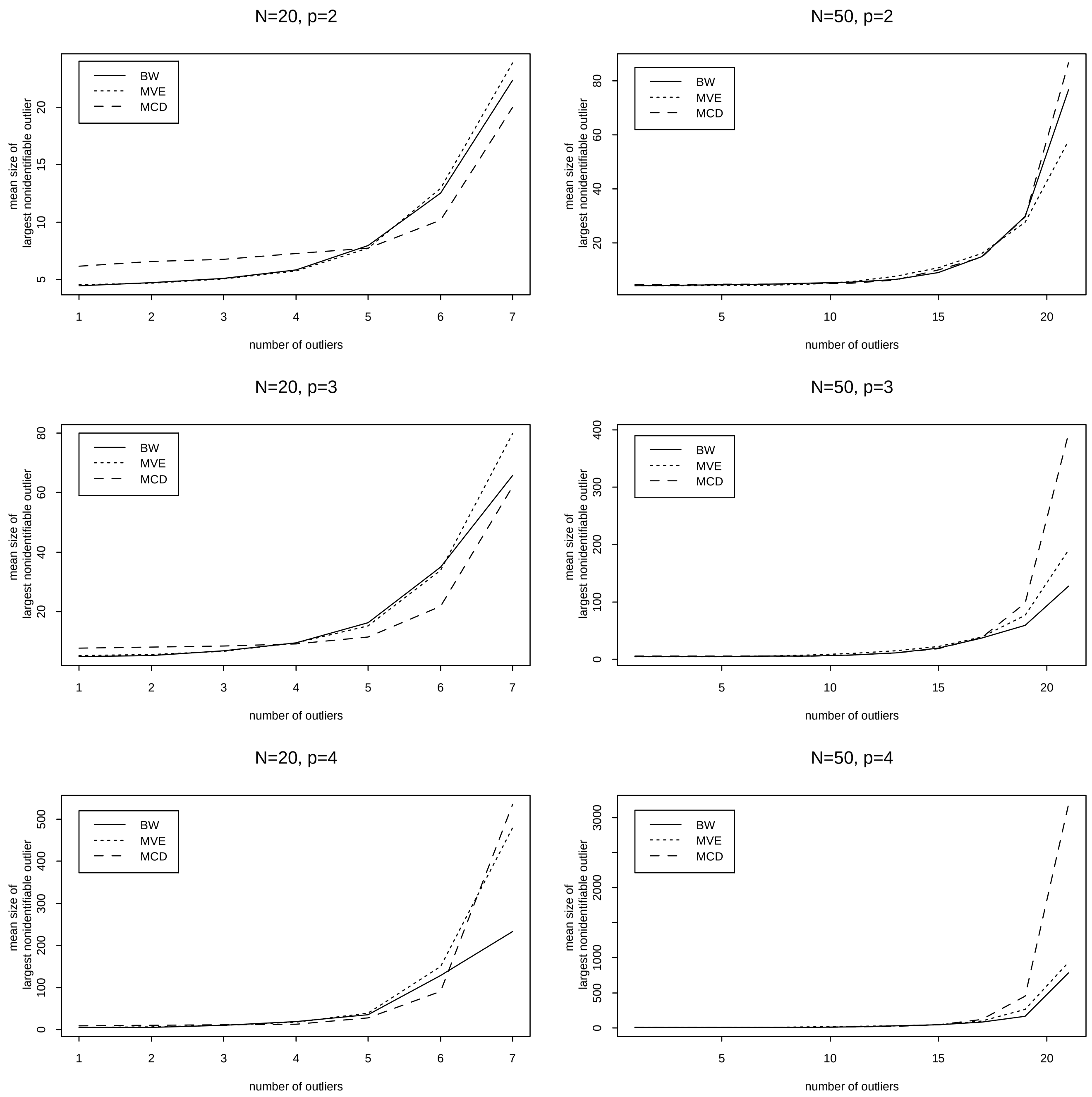

Fig. 1. Mean sizes of the largest nonidentifiable outliers

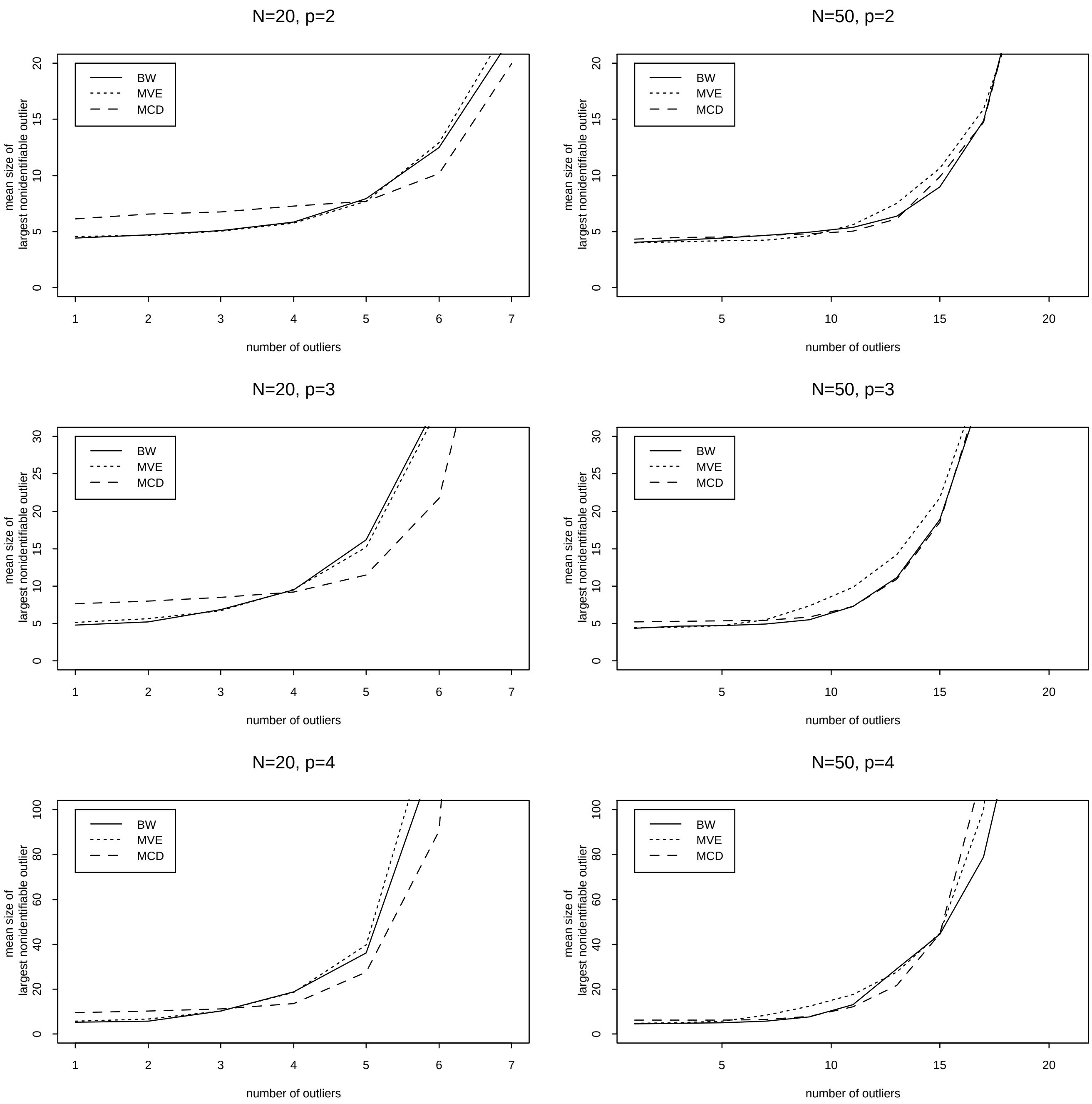

Fig. 2. Mean sizes of the largest nonidentifiable outliers: details

Acknowledgements

The authors thank Laurie Davies for his helpful discussion on an earlier version of this paper. The financial support of the Deutsche Forschungsgemeinschaft (SFB 475, "Reduction of complexity in multivariate data structures") is gratefully acknowledged.

References

Barnett, V., and Lewis, T., 1994. Outliers in Statistical Data, 3rd ed. Wiley, New York.

Beaton, A.E., and Tukey, J.W., 1974. The Fitting of Power Series, Meaning Polynomials, Illustrated on Band-Spectroscopic Data, *Technometrics* 16, 147–185.

Becker, C., and Gather, U., 1999. The Masking Breakdown Point of Multivariate Outlier Identification Rules, *J. Amer. Statist. Assoc.* 94, 947–955.

Croux, C., and Haesbroeck, G., 2000. Principal Component Analysis Based on Robust Estimators of the Covariance or Correlation Matrix: Influence Functions and Efficiencies. To appear in *Biometrika*.

Davies, P.L., 1987. Asymptotic Behaviour of S-Estimates of Multivariate Location Parameters and Dispersion Matrices. *Ann. Statist.* 15, 1269–1292.

Davies, P.L., and Gather, U., 1993. The Identification of Multiple Outliers, *J. Amer. Statist. Assoc.* 88, 782–792.

Donoho, D.L., 1982. Breakdown Properties of Multivariate Location Estimators. Ph.D. Qualifying Paper, Department of Statistics, Harvard University.

Donoho, D.L., and Huber, P.J., 1983. The Notion of Breakdown Point, in: P.J. Bickel, K.A. Doksum and J.L. Hodges (Eds.), A Festschrift for Erich Lehmann. Wadsworth, Belmont, CA, 157–184.

Gather, U., and Becker, C., 1997. Outlier Identification and Robust Methods, in: G.S. Maddala and C.R. Rao (Eds.), Handbook of Statistics, Vol. 15: Robust Inference. Elsevier, Amsterdam, 123–143.

Gather, U., and Hilker, T., 1997. A Note on Tyler's Modification of the MAD for the Stahel-Donoho Estimator, *Ann. Statist.* 25, 2024–2026.

Kent, J.T., and Tyler, D.E., 1996. Constrained M-estimation for Multivariate Location and Scatter, *Ann. Statist.* 24, 1346–1370.

Lopuhaä, H.P., 1989. On the Relation Between S-Estimators and M-Estimators of Multivariate Location and Covariance, *Ann. Statist.* 17, 1662–1683.

Lopuhaä, H.P., and Rousseeuw, P.J., 1991. Breakdown Points of Affine Equivariant Estimators of Multivariate Location and Covariance Matrices, *Ann. Statist.* 19, 229–248.

Maronna, R.A., and Yohai, V.J., 1995. The Behavior of the Stahel-Donoho Robust Multivariate Estimator, *J. Amer. Statist. Assoc.* 90, 330–341.

Rocke, D.M., 1996. Robustness Properties of S-Estimators of Multivariate Location and Shape in High Dimension, *Ann. Statist.* 24, 1327–1345.

Rousseeuw, P.J., 1985. Multivariate Estimation with High Breakdown Point, in: W. Grossmann, G. Pflug, I. Vincze and W. Wertz (Eds.), Mathematical Statistics and Applications. Reidel, Dordrecht, 283–297.

Rousseeuw, P.J., and van Driessen, K., 1999. A Fast Algorithm for the Minimum Covariance Determinant Estimator, *Technometrics* 41, 212–223.

Rousseeuw, P.J., and van Zomeren, B.C., 1990. Unmasking Multivariate Outliers and Leverage Points, *J. Amer. Statist. Assoc.* 85, 633–639.

Simonoff, J.S., 1987. The Breakdown and Influence Properties of Outlier Rejection-Plus-Mean Procedures, *Comm. Statist. A* 16, 1749–1760.

Stahel, W.A., 1981. Breakdown of Covariance Estimators. Research Report 31, Fachgruppe für Statistik, ETH Zürich.

Tyler, D.E., 1994. Finite Sample Breakdown Points of Projection Based Multivariate Location and Scatter Statistics, *Ann. Statist.* 22, 1024–1044.